DSSMATH

Utility Program for Mathematical Manipulation of HEC-DSS Data

User's Manual

January 1994

Hydrologic Engineering Center
U.S. Army Corps of Engineers
609 Second Street
Davis, California 95616-4687

(916) 756-1104

DSSMATH
Utility Program for
Mathematical Manipulation of HEC-DSS Data
User's Manual

Table of Contents

List of Figures

List of Tables

Appendices

PREFACE

This manual provides documentation for Program DSSMATH. It provides a detailed description of DSSMATH's capabilities and application.

Program DSSMATH provides capabilities for arithmetic computations and transformations of data stored in the Hydrologic Engineering Center's Data Storage System (HEC-DSS).

This document was prepared by Alfredo E. Montalvo, Technical Assistance Division, HEC. Art Pabst was Chief, Technical Assistance Division and Darryl W. Davis, Director, HEC, during preparation of this report.

DSSMATH

1. Purpose

DSSMATH enables mathematical manipulation of data stored in the Hydrologic Engineering Center's Data Storage System (HEC-DSS). The program provides capabilities for: arithmetic computations, transformations, such as stage to flow; screening; and estimation of missing or erroneous values. The program may be used in an automated batch environment for processing a real-time data stream, or it can be used interactively to perform ad hoc operations.

2. Description

DSSMATH supports regular, irregular-interval HEC-DSS time-series data and HEC-DSS paired function data. The standard memory configuration (UNIX and DOS Lahey Extended) can accommodate 10 time series records of up to 30000 values each, 5 paired-data records each with 30 dependent variables with a maximum of 500 values per record, and 15 scalars. Data may be accessed in several HEC-DSS files at the same time. Besides the DOS Lahey Extended implementation, a DOS version that runs under 640K of conventional memory is available. Its configuration is 5 time-series records consisting of 3180 values each and 15 paired functions consisting of 15 sets of dependent variables, with 300 values per paired function.

Time-series records and look-up tables are referenced in the various processing functions by user-assigned labels. The labels are one to six characters long and are assigned when the data is computed (CO command) or retrieved (GET command) from a HEC-DSS file. The HEC-DSS file name and data pathname are specified as parameters to the GET command (See Figure 1).

In either an interactive or in a batch setting, DSSMATH proceeds according to a sequence of instructions provided by the user. A typical sequence of instructions consists of a retrieval of data from the HEC-DSS file, a series of computations involving the data, and storage of results in a HEC-DSS file.

Computations in DSSMATH consist of arithmetic manipulations of time series data, such as multiplying a time series by a constant or another time series, and special functions that provide more complex and specific capabilities, such as rating table transformations.

Elementary screening functions flag data values internally to facilitate subsequent corrections. Replacements for flagged and missing values are provided by simple, linear estimation functions. Internal flags are used to assure that missing data do not erroneously affect computations. However, internal flags can not be read or written to HEC-DSS at present, and are not compatible with HEC-DSS Version 6 flag scheme.

The features of PREAD are provided in DSSMATH. PREAD facilitates efficient repetition of routine procedures. For example, a complex series of computations can be stored in a PREAD macro and started, for example, as a selection from a screen menu or a graphics tablet menu. (Ref: PREAD, Functions, Macros, Menus and Screens User Information)

3. Use

3.1 Running the Program

DSSMATH is initiated with: DSSMATH [parameter] ...

"parameter" has the form "key word=xxx" and may be one or more of the following specifications:

Key word	Default	Description
INPUT	terminal	Terminal or batch job input file
OUTPUT	terminal	Terminal or batch job output file
FUNFILE	MATHFUN	PREAD function file
MACFILE	MATHMAC	PREAD macro file
SCNFILE	GENSCN	PREAD screen definition file

In the interactive environment, the default output is the user's terminal display. In the batch environment, the default output is the job output file.

3.2 Use of Commands

The processing of time-series data with DSSMATH, either interactively or through a batch input file, is directed with commands. Commands are specified with two-character mnemonics in columns 1-2 of the input stream. The generalized form of the commands is: XX[.options] [parameter 1] [parameter 2] ....

Commands typically require parameters. Parameters supply specific information about the objects or entities to be processed or specifications to control the process.

Options are single characters used to specify alternative ways of processing.

Input lines beginning with two asterisks (**) are used for notes or comments and are simply echoed in the output.

A list of commands is presented in Table 1. The commands are described in detail in the appendices A and B.

Table 1 Commands

Command	Description
**	Comment.
CATALOG	Display a catalog of a DSS file.
CLEAR	Remove a variable data label from memory.
COMPUTE	Perform a computation.
DIAG	Toggles DSS diagnostic trace output on and off.
DPATH	Display a selective catalog of a DSS file.
FINISH	Terminate and exit the program.
GET	Retrieve data from a DSS file.
HELP	List commands.
OPEN	Open a DSS file.
PUT	Store data in a DSS file.
SD	Set data descriptions.
SHOW	Displays selected internal information about the data variable denoted by specified variable label.
SMISSING	Set missing value indicators.
SP	Set data pathname.
STATUS	Display key program variables states or values.
TABULATE	Tabulate values of time-series or paired data.
TIME	Set a time-window for time series data.
$CO	Resume processing subsequent to error.
$AB	Abort (stop) processing subsequent to error.

```
** Set a time window
TI T-2D T
**
** Get the "raw" data
GE STG=RAWDB:/SCIOTO/HIGH3/STAGE/01JUL1986/1HOUR/OBS/
**
** Get a look-up table for the station
GE TB=TABLEDB:/SCIOTO/HIGH3/STAGE-FLOW///USGS TABLE 6/
**
** Compute flows using the lookup table
COM FLW=RTABLE(STG,TB)
**
** Save the result
PUT FLW=MASTDB:/SCIOTO/HIGH3/FLOW/01JUL1986/1HOUR/COMP/
**
** Contingency checkpoint
$CONTINUE
**
** Clear program memories
CL  ALL
.
.
.
** Terminate
FI
```

Figure 1 Example Input

A typical command sequence retrieves a time series record, computes a dependent time series, and stores the result. An example input sequence that computes flow values from stage values is shown in Figure 1.

3.2.1 Data Management

3.2.1.1 Program Memory. The standard memory configuration can accommodate 10 time series records of up to 30000 values each, 5 paired-data records each with 30 dependent variables with a maximum of 500 values per record, and 15 scalars. The DOS version configuration is limited to 5 time-series records of 3180 values each, 15 paired-data records each with 15 dependent variables with 300 values per record, and 15 scalars.

Time series records may be regular or irregular and all data values carry a time and quality flag. The quality flags indicate four levels: 1) N for no flag; 2) E for estimated; 3) Q for questioned; and 4) M for undefined or missing. Time series records may be retrieved from files or may be computed in DSSMATH. However, currently, the data quality flags used in DSSMATH are not *the standard HEC-DSS quality flags and may not be read or stored to DSS.*

Paired data records include rating tables (ie., stage-flow) and polynomial coefficients. Paired-data records are usually retrieved from files, although at the present time only two compute functions," MATE and MRGP" manipulate a paired-data record.

Scalars are variables that are results of assignments or computations. Storage for scalars is provided in order to use the scalar result of a computation in a subsequent computation.

Each record or variable in the program memory is assigned an alphanumeric label as it is retrieved or computed. The label is then used to refer to that time-series in further processing. The label may be one to six characters long. It must begin with an alphabetic character but may contain numeric characters. A label should not be the same as any of the commands or function names.

The most recently used time window, pathname, and DSS file name are kept in the program's memory. A maximum of five DSS files can be kept open on a continuous basis. Once a sixth DSS file is requested, the first DSS file opened is automatically closed and the new DSS file requested is opened. The number of DSS files that can be open on a continuous basis is limited to three for the 640K DOS version.

3.2.1.2 Data Retrieval and Storage. The commands "GE" (get) and "PU" (put) are used, respectively, to retrieve and store time series and paired function data. The location of the data in DSS files is determined by both a file name and a pathname (See Figure 1). Data may be retrieved and stored in different files as desired. A warning message is printed if the retrieval exceeds any storage limits. Options are provided with the command "PU" to control overwriting of previously stored data.

The time window command "TI" controls the span of the data retrieved with "GE" and, therefore, defines the period of time for processing. The time window must be defined before any processing can be accomplished. The time window should be defined carefully, since arithmetic computations and several of the compute functions check for concurrence of the data.

When regular-interval time series data are retrieved, the time window is temporarily adjusted, if necessary, to include data belonging to exact interval boundaries. Options with the retrieval command "GE" provide retrievals outside the range of the time window to facilitate use of irregular-interval time series data.

3.2.1.3 Memory Management Functions. The status command ("ST") may be used to show memory contents, including all variable label names and types (time series, paired function, or scalar), the current time window, and the most recently used DSS pathname and DSS filename. "ST" options will also show the pathnames, units and types of time series and paired-function variables and values assigned to scalars. The command "CL" releases memory either globally or selectively, by label.

3.2.2 Computations. The COMPUTE command generates values of a dependent time series, paired function, or scalar from a simple arithmetic operation or a more complex function of one or more time series, paired functions, or scalars. If the dependent time series or paired function variable is not in memory it is created with undefined pathname, type, and units. However, the type and units may be assigned by the function being used.

The computation of a dependent time series may be subjected to a test based on a logical IF condition that compares values of two variables or a variable and a constant. The variables may be concurrent time series or scalars. If the IF condition is not satisfied for a particular time, then the corresponding value of the dependent variable is unchanged. If the value was not previously defined, it will remain undefined.

It is left to the user to provide appropriate parameters in computations: for example, the program will allow temperature values to be added to flows. However, arithmetic computations do require concurrent time series variables, and DSS data types or units are checked in certain compute functions. Usually, data types and units of a result must be specified explicitly as a separate processing step with the "SD" command.

3.2.2.1 Arithmetic Computations. An arithmetic computation of a time-series record consists of addition (+), subtraction (-), multiplication (*), division (/), or exponentiation (**) by a constant, previously defined scalar or time-series. Note that can not be used in arithmetic computaions. The following are examples:

```
COM Y=X+2     add 2 to time series "X"
COM Y=X/A     "A" may be a scalar or time-series
```

One of the independent variables in an arithmetic computation may be redefined (ie., the dependent variable may be the same as one of the independent variables). However, if a scalar and a time series variable are the independent variables, the dependent variable is a time series.

When arithmetic computations involve two independent time-series, the dependent variable will consist of the arithmetic combination of the concurrent successive values of the two series. "Concurrent successive values" means equal numbers of values with the same times. The "TS1" or "TS4" compute functions are useful for aligning two parallel time series vectors in time. The "TS1" function generates a regular-interval time series by interpolation at regular intervals. The "TS4" function generates a time series interpolated to the times of a pattern time series.

Two time series involved in an arithmetic computation may have different units and types. Units and types of results must be explicitly defined by the user with the "SD" command.

3.2.2.2 Functions. Computations may involve functions that operate upon one or more time-series records and result in a scalar or new time-series record. The functions are summarized below.

Table 2 Compute Functions

Name	*Description*
ABS	*Absolute function.*
ACC	*Running accumulation.*
AMREG	*Apply multiple linear regression equation.*
CMA	*Centered moving average smoothing.*
CONIC	*Conic interpolation based on elevation area table*
CORR	*Compute correlation coefficients.*
COS	*Cosine trigonometric function.*
COUNT	*Count the number of valid and missing data.*
DDT	*Differences per unit time.*
DECPAR	*Decaying basin wetness parameter.*
DIFF	*Successive differences.*
ESTLIN	*Estimate values for missing data.*
ESTPPT	*Estimate values for missing precipitation.*
FMA	*Forward moving average.*
GENTSR	*Generate a regular interval time series.*
INT	*Truncate to whole numbers.*
LAST	*Last valid value in a time series.*
LOG	*Natural log base "e".*
LOG10	*Log base 10.*
MATE	*Generate data pairs from two time series variables.*
MAX	*Maximum value in a time series.*
MEAN	*Mean value in a time series.*
MIN	*Minimum value in a time series.*
MREG	*Multiple linear regression coefficient funtion.*
MRG	*Merge two time series.*
MRGP	*Merge two paired data series.*
MUSK	*Muskingum routing function.*
NINT	*Round to nearest whole number.*
OLY	*Olympic smoothing.*

PERCON	*Period constants.*

Table 2 (continued)

Name	*Description*
POLY	*Polynomial transformation.*
POLY2	*Polynomial transformation with integral.*
PULS	*Modified Puls or Working R&D routing function.*
QAC	*Flow accumulator gage processor.*
RND	*Round off.*
RTABLE	*Rating table interpolation.*
RTABLR	*Reverse rating table interpolation.*
RTABL2	*Two variable rating table interpolation.*
SCRN1	*Screen for possible erroneous values based on max and min range.*
SCRN2	*Screen erroneous data values based on a forward moving average of max value.*
SDEV	*Compute the standard deviation of one independent variable.*
SELECT	*Extract time series data at unique time specification.*
SHIFT	*Shift adjustment.*
SIN	*Sine trigonometric function.*
SKEW	*Compute the skew coefficient of one independent variable.*
SQRT	*Square root function.*
SS	*Straddle Stagger routing function.*
SSW	*Willmington District Straddle Stagger routing function.*
TAN	*Tangent trigonometric function.*
TS1	*Interpolate data at regular time intervals.*
TS2	*Period averages at regular intervals.*
TS3	*Period minimums and maximums at regular intervals.*
TS4	*Interpolated data at irregular intervals.*
TSCYCL	*Time Series Cyclic Analysis.*
TSHIFT	*Shift time series in time.*
TSNAP	*Snap Irregualar times to nearest Regular period*
TTSR	*Transform time series to regular.*
TTSI	*Transform time series to irregular.*
1/X	*Inverse function.*

A detailed description and examples can be found in appendix B and D repectively.

In addition to the functions summarized above, DSSMATH allows the user who is proficient in FORTRAN, to create additional compute functions. See Appendix C for instructions on creating additional functions.

A detailed set of examples for each of the functions is available in Appendix D. Most of the examples are based on real applications developed at Corps of Engineers District offices.

Some computations result in a scalar value. Scalar values may also be assigned labels and used in subsequent processing, but cannot be retrieved or stored in DSS files.

3.2.3 Screening and Replacement

The screening and replacement capabilities of DSSMATH are implemented in the compute functions "SCRN1", "SCRN2", "ESTLIN", and "ESTPPT". Data values that exceed the specified limits are flagged internally to facilitate subsequent corrections with estimation functions or graphical editing. Note: Program DATCHK is also capable in performing this type of screening and is substantially more powerful than DSSMATH.

3.2.4 Tabulation and Show commands

Tabulation capabilities of DSSMATH are nearly identical with those of the program DSPLAY. Time series and paired function data may be tabulated with the "TA" command. The "SHOW" command can also be used to tabulate the variables and has the additional capability of displaying scalar variables.

3.2.5 Miscellaneous Commands

The DSS file catalog command "CA" is used for listing the pathnames of records in a DSS file. Data variable pathnames and descriptions (eg., units) can be specified using the "SP" and "SD" commands, respectively. On-line help is available through the "HE" command. The finish command, "FI", terminates processing.

3.3 Contingency Processing

Certain error conditions can render subsequent processing inappropriate. For example, if data cannot be retrieved, it cannot be processed. These errors, which simply result in warnings in the interactive environment, may be processed with recovery procedures when input are taken from an input file or a PREAD macro by specifying recovery checkpoints in the input.

An input line beginning with a dollar sign ($) is an error recovery checkpoint. "$CONTINUE" indicates where to resume processing when an error is encountered. "$ABORT" indicates the program is to stop if an error has been encountered and no "$CONTINUE" was found. If no recovery checkpoints are found, processing stops.